Manifold-Aware Spectral Compaction: A Graph Signal Processing Perspective on Online Gaussian Reduction for 3DGS SLAM

Jiayi Tian[1], Jiaze Wang[2], Tian Xia[1], Wenzhe Zhao[1], and Pengju Ren[1*]

[1] State Key Laboratory of Human-Machine Hybrid Augmented Intelligence, Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, China
[2] School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, China
{tian_xia, wenzhe, pengjuren}@xjtu.edu.cn
tianreg@stu.xjtu.edu.cn 24b951071@stu.hit.edu.cn

Abstract. 3D Gaussian Splatting (3DGS) has established itself as a premier technique for high-fidelity radiance field rendering. However, its application in SLAM tasks is frequently hindered by the explosive growth of redundant Gaussian primitives, which imposes prohibitive memory and rendering overhead. Existing compaction strategies predominantly rely on heuristic pruning based on importance scores; such methods often fail to provide global fidelity guarantees and can disrupt the optimization continuity required for online tracking. To address these challenges, we introduce **M**anifold-**A**ware **S**pectral **C**ompaction (**MASC**), a framework that reformulates map reduction as a continuous signal reconstruction task mapped onto a Riemannian manifold. We first construct a topology-aware graph representation using Symmetrized Kullback-Leibler divergence to precisely capture the statistical connectivity between primitives. To ensure real-time efficiency, we implement an incremental Nyström spectral embedding strategy to project the Gaussian map onto a low-dimensional spectral subspace. We further derive a mathematically rigorous, closed-form solution for primitive aggregation leveraging Bures-Wasserstein barycenters, which theoretically ensures the conservation of local radiance field moments. Extensive evaluations on the Replica and TUM RGB-D datasets demonstrate that MASC achieves a **2.8× reduction** in GPU memory and a **2.4× speedup** in rendering throughput. Remarkably, our method maintains or even enhances tracking accuracy while eliminating approximately **60%** of redundant primitives, achieving an ATE of **0.98 cm** on the TUM dataset. These results substantiate that MASC provides an efficient and principled plug-and-play pathway for scalable, long-term neural rendering SLAM.

Keywords: 3D Gaussian Splatting · SLAM · Model Compaction · Optimal Transport · Spectral Embedding

* Corresponding author. This work was supported in part by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China under Grant JYB2025XDXM504, and the National Natural Science Foundation of China under Grants 62302381 and 52441602.

1 Introduction

3D Gaussian Splatting (3DGS) has established itself as a transformative representation for dense SLAM, enabling real-time and photo-realistic rendering through differentiable anisotropic Gaussian primitives [7, 35, 41]. Compared to traditional voxel- or mesh-based mapping paradigms, 3DGS provides superior rendering fidelity and continuous spatial modeling [29, 40, 43]. These characteristics make it a premier candidate for applications in augmented reality, autonomous robotics, and embodied artificial intelligence.

However, the integration of 3DGS into online SLAM pipelines introduces a fundamental scalability bottleneck [3, 18, 27]. As newly observed views continuously trigger densification to minimize photometric residuals, the map often suffers from an uncontrolled proliferation of primitives [10, 25, 36, 45]. This phenomenon leads to significant memory consumption and the accumulation of rendering latency [8, 24, 33, 44]. In scenarios involving long trajectories, the memory footprint scales almost linearly with time, which limits the deployment of such systems on resource-constrained edge platforms.

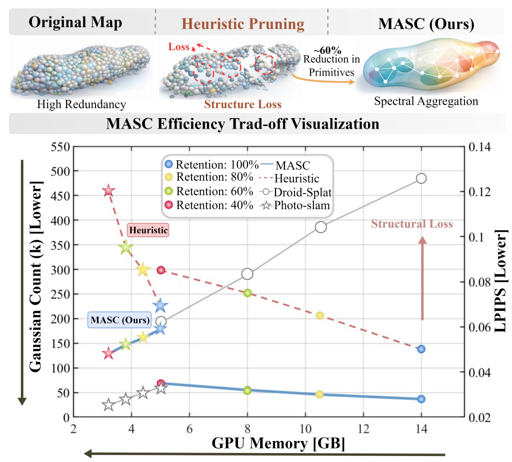

Fig. 1: Motivation of the proposed MASC. Top: Heuristic pruning leads to significant "Structure Loss" and fragmentation, while MASC ensures smooth, manifold-consistent surfaces via spectral aggregation. Bottom: Efficiency trade-off for *DROID-Splat* and *Photo-SLAM* backbones. Unlike heuristic methods that spike LPIPS (red arrow), MASC achieves ∼60% primitive reduction and lower GPU Memory while maintaining high reconstruction fidelity.

Existing compaction strategies predominantly rely on offline post-processing and typically assume access to a fully converged static map [24, 39, 49]. These paradigms are often unsuitable for online SLAM for three primary reasons. First, they require global map availability, which precludes real-time operation. Second, they frequently rely on heuristic importance scores or Euclidean spatial grouping, which ignores the intrinsic statistical overlap between Gaussians. Third, abrupt pruning or merging can disrupt gradient continuity, thereby inducing tracking drift and surface fragmentation.

As illustrated in Fig. 1, while heuristic pruning effectively reduces map density, it inevitably induces severe structural loss and surface discontinuity, as evidenced by the sharp spike in LPIPS error. This observation drives our contention that the redundancy in 3DGS-SLAM is not merely a spatial over-density problem but is fundamentally a spectral redundancy phenomenon. As the map converges, neighboring Gaussians representing the same physical surface exhibit near-identical attributes. From a Graph Signal Processing (GSP) perspective, the map signal energy concentrates overwhelmingly in the low-frequency subspace of the graph Laplacian. Consequently, the system over-allocates primitives to represent smooth signal components that could be described more efficiently.

To address these challenges, we propose **M**anifold-**A**ware **S**pectral **C**ompaction (**MASC**), a framework that reformulates map reduction as a continuous signal reconstruction task mapped onto a Riemannian manifold. Rather than relying on heuristic pruning, we reinterpret map compression as a signal reconstruction problem in the spectral domain. By aggregating Gaussians across the spectrum, we selectively collapse redundant low-frequency components into compact Super-Gaussians while preserving the high-frequency modes that characterize geometric discontinuities. Leveraging the Bures-Wasserstein barycenter, our mechanism ensures the conservation of local radiance field moments. To the best of our knowledge, this work represents the first attempt to address online 3DGS-SLAM compaction from the perspective of Manifold-Aware Spectral Compaction, providing a principled and plug-and-play Graph Signal Processing interpretation for Gaussian redundancy.

Our main contributions are summarized as follows:

- **Statistical Topology Modeling:** We introduce a topology-aware graph construction based on Symmetrized Kullback-Leibler (SKL) divergence, which captures manifold connectivity through statistical overlap rather than simple Euclidean proximity.
- **Spectral Redundancy Analysis:** We reveal that mature 3DGS-SLAM maps exhibit vanishing Graph Dirichlet Energy, providing a principled GSP interpretation of Gaussian redundancy within the low-frequency spectral subspace.
- **Incremental Spectral Embedding:** We develop an incremental spectral embedding strategy leveraging Nyström approximation. This enables scalable eigenspace analysis with $\mathcal{O}(nm)$ complexity to support real-time SLAM operations.
- **Moment-Preserving Compaction:** We formulate an adaptive objective function that jointly minimizes spectral reconstruction error and geometric dilation. By utilizing Wasserstein barycenters, we ensure structural stability and gradient continuity during online map compaction.

2 Related Work

2.1 Geometric Redundancy and Primitive Compaction in 3DGS

3D Gaussian Splatting (3DGS) [11,23,31] has established itself as a high-performance alternative to coordinate-based neural radiance fields [1, 2, 50], offering exceptional rendering speed through anisotropic Gaussian primitives. Despite these advantages, the explicit nature of 3DGS often results in a massive number of primitives, imposing significant memory and computational burdens [20,26]. Existing literature has explored various compaction strategies, which generally fall into two categories: heuristic pruning and structural aggregation.

Heuristic pruning methods identify and remove low-impact primitives based on predefined importance scores. Representative works [8,9,12] employ criteria like accumulated ray contribution, opacity, or Hessian sensitivity to rank Gaussians. While effective in reducing map size, these methods evaluate primitives independently. Such a localized perspective often neglects the global structural consistency of the scene manifold, leading to potential holes in the reconstructed geometry.

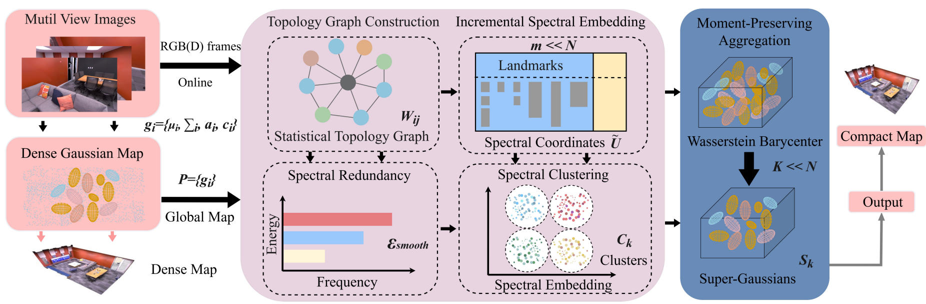

Fig. 2: Overview of the proposed **MASC** framework. Given online RGB(D) frames, a dense Gaussian map is incrementally constructed. We first model the global map as a Statistical Topology Graph based on the statistical overlap of primitives. Then, an **Incremental Spectral Embedding** strategy is employed to obtain low-dimensional spectral coordinates $\tilde{U}$, where $m \ll N$ landmark nodes are selected via farthest point sampling. By analyzing the Spectral Redundancy, we identify that map signal energy is concentrated in the low-frequency domain. Redundant primitives are subsequently grouped through Spectral Clustering in the embedding space. Finally, we leverage a Moment-Preserving Aggregation mechanism based on the Wasserstein barycenter to collapse these clusters into compact Super-Gaussians (S_k), resulting in a high-fidelity compact map where $K \ll N$.

Structural aggregation approaches, on the other hand, attempt to merge redundant primitives into more compact representations. For instance, several studies [4, 38, 42] leverage Gaussian mixture reduction to provide a principled global surrogate for the primitive set. However, most existing aggregation frameworks are designed for the offline post-processing of static scenes. These paradigms assume global map availability and are not directly applicable to the incremental and non-stationary nature of online SLAM systems.

2.2 Online Map Maintenance and Topology-Preserving SLAM

Extending 3DGS to the SLAM domain has enabled simultaneous camera tracking and dense reconstruction with impressive efficiency. Pioneering systems [15, 17, 46] utilize 3DGS as the underlying map representation. These methods typically employ the standard densification mechanism of 3DGS, where new primitives are introduced to minimize photometric residuals as new views are observed. While effective for short sequences, this incremental densification causes the primitive count to scale almost linearly with trajectory length, leading to prohibitive GPU memory consumption [14, 19].

To manage map growth, existing SLAM systems often rely on simplistic map maintenance heuristics, such as visibility-based filtering or threshold-based pruning of low-opacity Gaussians [21, 22, 28]. These independent-primitive strategies frequently ignore the underlying manifold topology. Abruptly deleting Gaussians can disrupt geometric continuity, potentially leading to surface fragmentation and instability in the pose optimization loop.

In contrast to these heuristic-based maintenance strategies, our MASC framework bridges the gap between principled global aggregation and real-time SLAM requirements. By reinterpreting map reduction as a spectral signal reconstruction task, we achieve a topology-preserving reduction that operates seamlessly within the dynamic mapping loop. This approach ensures that the fundamental structural integrity of the map is maintained, thereby preventing the tracking degradation commonly associated with abrupt map pruning.

3 Methodology

We reinterpret the reduction of 3DGS maps as a continuous signal reconstruction task defined on a topology-aware Riemannian manifold, departing from traditional discrete and heuristic pruning paradigms. The overall pipeline is illustrated in Fig. 2. Our methodology begins by constructing a statistical topology graph that captures the intrinsic connectivity between Gaussian primitives (Sec. 3.2). We then characterize the structural redundancy within this representation through the lens of spectral graph theory and Graph Dirichlet Energy (Sec. 3.3). To enable real-time execution during online SLAM, we implement an incremental Nyström strategy for scalable spectral embedding (Sec. 3.4). Finally, we introduce a principled aggregation mechanism that leverages Wasserstein barycenters to generate compact Super-Gaussians, ensuring the strict preservation of radiance field moments (Sec. 3.5).

3.1 Spectral Compaction on Manifolds

We provide a theoretical interpretation of our framework under a manifold assumption commonly adopted in spectral graph theory.

We assume that the Gaussian centers $\{\mu_i\}_{i=1}^{N}$ are sampled from a smooth, compact, $d_{\mathcal{M}}$ dimensional Riemannian manifold $\mathcal{M} \subset \mathbb{R}^3$, embedded in the ambient Euclidean space. This assumption is consistent with the observation that 3DGS primitives densely populate locally smooth surface patches in mature SLAM maps.

Let $\mathbf{L}_N$ denote the graph Laplacian constructed using a symmetric kernel with bandwidth σ. Under standard regularity conditions and as $N \to \infty$, the properly normalized graph Laplacian converges to the Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ on the manifold $\mathcal{M}$:

$$\mathbf{L}_N f \longrightarrow c\,\Delta_{\mathcal{M}} f, \tag{1}$$

where c is a constant depending on the kernel choice.

Dirichlet Energy as Discrete Manifold Energy. The graph Dirichlet energy:

$$\mathcal{E}_{smooth}(\mathbf{F}) = \mathrm{tr}(\mathbf{F}^T \mathbf{L}_N \mathbf{F}) \tag{2}$$

serves as a discrete approximation of the continuous manifold energy

$$\mathcal{E}_{\mathcal{M}}(f) = \int_{\mathcal{M}} \|\nabla_{\mathcal{M}} f(x)\|^2 \, d\mu(x), \tag{3}$$

where $\nabla_{\mathcal{M}}$ denotes the intrinsic gradient on $\mathcal{M}$.

Spectral Compaction on the Manifold. Minimizing spectral reconstruction error in our framework therefore corresponds to preserving the low-frequency eigenspaces of the Laplace–Beltrami operator. Since these eigenspaces encode large-scale geometric structure of the manifold, our compaction strategy is intrinsically manifold-aware: it suppresses redundant low-variation components while retaining high-frequency modes that characterize geometric discontinuities.

3.2 Statistical Topology Graph Construction

Let the incrementally built global map be denoted as $\mathcal{P} = \{g_i\}_{i=1}^N$, comprising N Gaussian primitives. Each individual primitive g_i is fully parameterized by its spatial mean $\mu_i \in \mathbb{R}^3$, covariance matrix $\Sigma_i \in \mathbb{R}^{3\times 3}$, opacity scalar α_i, and spherical harmonics (SH) coefficient vector $\mathbf{c}_i$. We structurally model $\mathcal{P}$ as a weighted undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$.

Unlike conventional compaction methods that heavily rely on Euclidean distance-based KD-trees or spatial hashing, we argue that Euclidean proximity frequently fails to encapsulate true topological connectivity, such as disjoint physical surfaces located on opposite sides of a thin structural barrier. Therefore, we define the graph edge weights based on the underlying statistical overlap of the Gaussian Probability Density Functions (PDFs). The physical interaction between any two Gaussians g_i and g_j is precisely quantified by the Symmetrized Kullback-Leibler (SKL) Divergence:

$$\begin{aligned} D_{\mathrm{SKL}}(\mathcal{N}_i \| \mathcal{N}_j) = \frac{1}{2}\Big(&\mathrm{tr}(\Sigma_j^{-1}\Sigma_i) + \mathrm{tr}(\Sigma_i^{-1}\Sigma_j) \\ &+ (\mu_j - \mu_i)^\top (\Sigma_i^{-1} + \Sigma_j^{-1})(\mu_j - \mu_i) - 2d\Big) \end{aligned} \tag{4}$$

where $D_{SKL}(\cdot\|\cdot)$ denotes the symmetrized divergence between the multivariate normal distributions $\mathcal{N}_i$ and $\mathcal{N}_j$ corresponding to the i-th and j-th primitives, respectively, $\mathrm{tr}(\cdot)$ represents the trace operator, and d explicitly defines the dimensionality of the ambient space, which is set to $d = 3$ for 3DGS.

Building upon this divergence, the adjacency matrix weight W_{ij} representing the edge between node i and node j is formulated as a truncated Gaussian kernel that jointly evaluates geometric and appearance similarity:

$$W_{ij} = \begin{cases} \exp\left(-\frac{D_{SKL}(\mathcal{N}_i\|\mathcal{N}_j)}{\sigma_{geo}^2} - \frac{\|\mathbf{c}_i - \mathbf{c}_j\|_2^2}{\sigma_{app}^2}\right), & \text{if } j \in \mathrm{k}_{nn}(i) \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

In this formulation, the bandwidth hyper-parameters σ_{geo} and σ_{app} dictate the kernel's sensitivity to geometric divergence and appearance discrepancies, respectively. To rigorously enforce graph sparsity and ensure computational tractability, this weight is exclusively computed if node j resides within the k-nearest spatial neighbors of node i, denoted by the set $\mathrm{k}_{nn}(i)$; otherwise, the interaction weight W_{ij} is explicitly forced to zero.

3.3 Spectral Redundancy and Dirichlet Energy

We treat the aggregated attributes of the Gaussian primitives as a continuous, multi-channel signal $\mathbf{F} \in \mathbb{R}^{N \times D}$ residing entirely on the graph vertices. The topological structure of the map is algebraically encoded by the Graph Laplacian $\mathbf{L}_N = \mathbf{D} - \mathbf{W}$, with $\mathbf{D}$ serving as the diagonal degree matrix whose individual elements are computed as $\mathbf{D}_{ii} = \sum_j W_{ij}$. To mathematically capture the structural redundancy inherently generated during continuous view synthesis, we introduce the Graph Dirichlet Energy to measure the overarching smoothness of the map signal:

$$\mathcal{E}_{smooth}(\mathbf{F}) = \mathrm{tr}(\mathbf{F}^T \mathbf{L}_N \mathbf{F}) = \frac{1}{2} \sum_{i,j} W_{ij} \|\mathbf{f}_i - \mathbf{f}_j\|_2^2 \tag{6}$$

where $\mathcal{E}_{smooth}(\mathbf{F})$ quantifies the total variation of the signal across the graph topology, and $\mathbf{f}_i, \mathbf{f}_j \in \mathbb{R}^D$ represent the concatenated feature vectors at nodes i and j, where $\mathbf{f}_i^\top$ corresponds to the i-th row of $\mathbf{F}$. In a mature, well-converged 3DGS-SLAM map, densely packed neighboring primitives describing the exact same continuous surface patch inevitably exhibit exceptionally high attribute similarity (i.e., $\mathbf{f}_i \approx \mathbf{f}_j$). This phenomenon organically drives the Dirichlet energy $\mathcal{E}_{smooth}(\mathbf{F}) \to 0$, indicating that the operational signal energy is heavily concentrated within the low-frequency spectral subspace spanned by the Laplacian matrix.

3.4 Scalable Incremental Spectral Embedding

Executing a full eigendecomposition of the Laplacian matrix incurs an $\mathcal{O}(N^3)$ computational complexity, a prohibitive cost that entirely precludes its direct application in online SLAM pipelines. To seamlessly bypass this bottleneck, we implement a block-wise Nyström approximation strategy. During dynamic map maintenance, we systematically partition the active map into an landmark set of landmark nodes $\mathcal{A}$ with cardinality $|\mathcal{A}| = m$, alongside the remaining supplementary nodes $\mathcal{B}$ where $|\mathcal{B}| = N - m$ and $m \ll N$. The landmarks are strictly selected via Farthest Point Sampling (FPS) to ensure maximal spatial and topological coverage. Consequently, the global affinity matrix $\mathbf{W}$ is approximated as:

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{\mathcal{A}\mathcal{A}} & \mathbf{W}_{\mathcal{A}\mathcal{B}} \\ \mathbf{W}_{\mathcal{B}\mathcal{A}} & \mathbf{W}_{\mathcal{B}\mathcal{B}} \end{bmatrix} \approx \mathbf{C}\mathbf{W}_{\mathcal{A}\mathcal{A}}^{\dagger}\mathbf{C}^T, \quad \text{where } \mathbf{C} = \begin{bmatrix} \mathbf{W}_{\mathcal{A}\mathcal{A}} \\ \mathbf{W}_{\mathcal{B}\mathcal{A}} \end{bmatrix} \tag{7}$$

Here, the full matrix is elegantly reconstructed using the $m \times m$ landmark affinity sub-matrix $\mathbf{W}_{\mathcal{A}\mathcal{A}}$, its Moore-Penrose pseudo-inverse denoted by $\mathbf{W}_{\mathcal{A}\mathcal{A}}^{\dagger}$, and the concatenated block matrix $\mathbf{C} \in \mathbb{R}^{N \times m}$ which explicitly houses the cross-affinities $\mathbf{W}_{\mathcal{A}\mathcal{B}}$ and $\mathbf{W}_{\mathcal{B}\mathcal{A}}$ connecting landmarks to the rest of the graph. Moreover, when a new keyframe is introduced, the entire graph does not need to be recomputed; only the corresponding rows in the block matrix $\mathbf{C}$ associated with the newly added nodes require updating.

Let the eigendecomposition (EVD) of the small landmark matrix be computed as $\mathbf{W}_{\mathcal{A}\mathcal{A}} = \mathbf{U}_{\mathcal{A}}\mathbf{\Lambda}_{\mathcal{A}}\mathbf{U}_{\mathcal{A}}^T$, where $\mathbf{U}_{\mathcal{A}}$ and $\mathbf{\Lambda}_{\mathcal{A}}$ represent its eigenvector and diagonal eigenvalue matrices, respectively. Leveraging this decomposition, the $N \times m$ matrix containing the approximated eigenvectors for the entire global graph, denoted as $\tilde{\mathbf{U}}$, can be analytically derived via:

$$\tilde{\mathbf{U}} = \mathbf{C}\mathbf{U}_{\mathcal{A}}\mathbf{\Lambda}_{\mathcal{A}}^{-1} \tag{8}$$

Following the embedding, K-Means clustering is performed on the rows of $\tilde{\mathbf{U}}$ to group Gaussians with similar spectral coordinates into clusters of structural redundancy for subsequent aggregation.

3.5 Moment-Preserving Super-Gaussian Aggregation

Once a distinct spectral cluster $\mathcal{C}_k$ is isolated, our subsequent objective is to seamlessly merge its constituent primitive members $\{g_i\}_{i \in \mathcal{C}_k}$ into a unified, high-fidelity "Super-Gaussian" S_k. Standard arithmetic averaging or linear interpolation of covariance matrices inherently destroys the geometric volume and rotational anisotropy of the original primitives. To strictly preserve these attributes, we compute the exact Wasserstein Barycenter of the constituent Gaussian distributions. The optimal covariance matrix Σ_{S_k} for the aggregated Super-Gaussian is formulated to minimize the sum of weighted squared 2-Wasserstein distances:

$$\Sigma_{S_k} = \arg\min_{\Sigma} \sum_{i \in \mathcal{C}_k} w_i d_W^2(\mathcal{N}(0, \Sigma), \mathcal{N}(0, \Sigma_i)) \tag{9}$$

where $d_W^2(\cdot, \cdot)$ denotes the squared 2-Wasserstein distance between two respective distributions, and $w_i = \alpha_i / \sum_{j \in \mathcal{C}_k} \alpha_j$ provides a normalized, opacity-based contribution weight for the i-th Gaussian within the cluster. This non-convex optimization is robustly solved via a closed-form fixed-point iteration scheme:

$$\Sigma_{S_k}^{(t+1)} = \left(\Sigma_{S_k}^{(t)}\right)^{-1/2} \left(\sum_{i \in \mathcal{C}_k} w_i \left(\left(\Sigma_{S_k}^{(t)}\right)^{1/2} \Sigma_i \left(\Sigma_{S_k}^{(t)}\right)^{1/2} \right)^{1/2} \right)^2 \left(\Sigma_{S_k}^{(t)}\right)^{-1/2} \tag{10}$$

where t designates the current iteration step. Conversely, for the linear attributes encompassing the spatial position μ_{S_k} and the appearance vector $\mathbf{c}_{S_k}$, we directly apply a weighted Graph Low-Pass Filter defined as:

$$\mu_{S_k} = \sum_{i \in \mathcal{C}_k} w_i \mu_i, \quad \mathbf{c}_{S_k} = \sum_{i \in \mathcal{C}_k} w_i \mathbf{c}_i \tag{11}$$

This sophisticated aggregation paradigm theoretically guarantees the strict preservation of the first and second-order moments of the local radiance field, yielding a compact representation without compromising rendering fidelity.

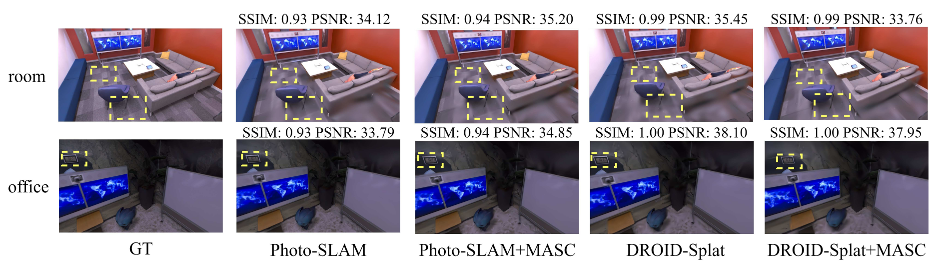

Fig. 3: Qualitative comparison on the Replica dataset. Even after a ∼60% reduction in primitive count, our **MASC** framework achieves high-fidelity reconstruction comparable to the uncompressed backbones. As highlighted in the yellow boxes, MASC effectively preserves sharp geometric details and structural integrity, whereas baseline methods often suffer from over-smoothing or obvious artifacts in complex regions.

3.6 Optimization Objective

Ultimately, the macroscopic map retention ratio is not a manually hard-coded heuristic. We formally define this ratio as $\rho_k = K/|\mathcal{P}|$, where $K = |\{\mathcal{C}_k\}|$ denotes the number of aggregated Super-Gaussians and $|\mathcal{P}|$ is the cardinality of the original primitive set. Instead of fixing ρ_k artificially, our framework adaptively steers it through an objective function designed to minimize the spectral reconstruction error while strictly adhering to a physical density constraint:

$$\min_{\{\mathcal{C}_k\}} \sum_{k=1}^{K} \sum_{i \in \mathcal{C}_k} \|\mathbf{u}_i - \mathbf{z}_k\|_2^2 + \lambda \sum_{k=1}^{K} \mathrm{Vol}(\Sigma_{S_k}) \tag{12}$$

Table 1: Quantitative results on the Replica dataset. We evaluate our method against various baselines under both Monocular and RGB-D modalities. Notably, the results of our method are reported after compressing the total number of 3D Gaussians by approximately 60%. We mark the **best** and second best results. The rows corresponding to our proposed methods are highlighted in yellow. NICE-SLAM* indicates that depth supervision is disabled. "-" denotes that the system does not support the metric or fails to track.

Input Method	Localization (cm)		Mapping Quality			Resource Efficiency (RTX 4090)		
	RMSE ↓	STD ↓	PSNR ↑	SSIM ↑	LPIPS ↓	T-FPS ↑	R-FPS ↑	GPU Memory ↓
Monocular / RGB Input								
NICE-SLAM* [51]	99.94	35.34	16.31	0.720	0.439	2.38	0.94	12 GB
Orbeez-SLAM [5]	-	-	23.25	0.790	0.336	**49.20**	1.03	6 GB
Go-SLAM [48]	71.05	24.59	21.17	0.703	0.421	25.37	0.82	22 GB
MonoGS [27]	14.54	-	31.22	0.910	0.210	-	-	-
Splat-SLAM [32]	0.35	0.54	36.22	0.950	0.060			
Photo-SLAM [15]	1.09	0.89	33.30	0.926	0.078	41.65	911.3	6 GB
DROID-Splat [13]	**0.27**	**0.16**	**39.47**	**1.000**	**0.030**	30.00	145.0	6 GB
MASC+Photo-SLAM (Ours)	0.35	0.88	34.15	0.935	0.065	42.00	**1150.0**	3.5 GB
MASC+DROID-Splat (Ours)	**0.27**	**0.16**	39.20	**1.000**	0.031	30.00	350.0	**2.3 GB**
RGB-D Input								
BundleFusion [6]	1.60	0.97	23.84	0.822	0.197	8.63	-	5 GB
NICE-SLAM [51]	2.35	1.59	26.16	0.832	0.232	2.33	0.61	12 GB
Orbeez-SLAM [5]	0.88	0.56	32.52	0.916	0.112	41.33	1.40	6 GB
ESLAM [16]	0.56	0.27	30.59	0.866	0.162	6.69	2.63	21 GB
Co-SLAM [37]	1.15	0.60	30.25	0.864	0.175	14.58	3.75	4 GB
Go-SLAM [48]	0.57	0.22	24.16	0.766	0.352	19.44	0.44	24 GB
Point-SLAM [30]	0.59	0.24	34.63	0.921	0.083	0.35	0.51	24 GB
SplaTaM [17]	3.40	-	22.00	0.862	0.230	19.40	0.44	24 GB
Gaussian-SLAM [46]	0.31	-	**42.08**	0.990	**0.018**	-	-	-
Photo-SLAM [15]	0.60	0.29	34.96	0.940	0.059	42.49	1084.0	5 GB
Splat-SLAM [32]	0.35	0.24	36.45	0.949	0.060	-	-	-
DROID-Splat [13]	**0.29**	0.17	39.66	**1.000**	0.030	30.00	145.0	14.0 GB
MASC+Photo-SLAM (Ours)	0.30	0.28	36.39	0.950	0.048	**65.00**	**1103.0**	3.2 GB
MASC+DROID-Splat (Ours)	**0.29**	**0.13**	39.16	**1.000**	0.037	30.00	350.0	**5.0 GB**

where $\mathbf{u}_i$ denotes the spectral embedding vector of node i (extracted directly as a row from $\tilde{\mathbf{U}}$), and $\mathbf{z}_k$ represents the centroid of cluster $\mathcal{C}_k$ within the corresponding spectral space. The regularization parameter λ critically dictates the trade-off between reconstruction fidelity and geometric sparsity, thereby implicitly governing the final retention ratio ρ_k. Concurrently, the term $\mathrm{Vol}(\Sigma_{S_k}) \propto \sqrt{\det(\Sigma_{S_k})}$ intuitively computes the volumetric span of the newly formed Super-Gaussian, explicitly penalizing the generation of excessively dilated primitives that could induce rendering artifacts.

4 Experiments

4.1 Implementation and Experimental Setup

Implementation Details. We developed MASC as a high-performance, modular framework with CUDA-accelerated kernels to ensure seamless integration into diverse SLAM architectures. This design allows MASC to function as a plug-and-play component across different implementation environments. The compression is triggered dynamically based on the mapping iterations. For the Nyström approximation, the landmark sampling rate M is set between 1%–5% of the total primitives [13]. All experiments were conducted on a desktop workstation equipped with an NVIDIA RTX 4090 (24 GB GPU), an Intel Core i9-13900K CPU, and 64 GB RAM. To showcase the efficiency, we also provide comparative analysis regarding memory footprints.

Datasets and Baselines. We conduct comprehensive evaluations on the **Replica** [34] and **TUM-RGBD** [15] datasets to validate MASC across diverse environments. **Replica** provides high-fidelity synthetic indoor scenes, while **TUM-RGBD** introduces real-world challenges. The baselines include SOTA 3DGS-based systems, as well as volume-rendering-based systems.

Fig. 4: Visual quality evaluation on the TUM RGB-D dataset. Despite the presence of rapid camera motion and motion blur, MASC maintains sharp textures and structural consistency with only $\sim$40% of the original Gaussian primitives. The yellow boxes illustrate that our spectral aggregation effectively regularizes the noisy Gaussian distribution, mitigating rendering artifacts caused by chaotic dynamics while preserving fine-grained details.

Metrics. We estimate localization accuracy using the Absolute Trajectory Error (ATE) RMSE [cm]. For rendering quality, we report PSNR, SSIM, and LPIPS [17] on rendered images against the ground truth. Crucially, to evaluate the system's resource efficiency, we report tracking FPS, rendering FPS, peak GPU memory usage, and final map storage size. To mitigate the nondeterministic nature of multi-threaded optimization, all reported results are averaged over five independent runs.

Table 2: Quantitative evaluation on the TUM RGB-D dataset. We report the tracking accuracy (ATE RMSE [cm] ↓) and rendering quality (PSNR ↑, SSIM ↑, LPIPS ↓) across monocular and RGB-D configurations. Consistent with our Replica experiments, the performance of our method is evaluated based on a ∼60% reduction in Gaussian primitives. The **best** and second best results are highlighted, and the rows presenting our proposed methods are shaded in yellow.

Input Method	fr1/desk				fr2/xyz				fr3/office			
	ATE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	ATE ↓	PSNR ↑	SSIM ↑	LPIPS ↓	ATE ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Monocular / RGB Input												
GO-SLAM [48]	2.119	15.794	0.531	0.538	31.788	16.12	0.53	0.42	26.80	16.50	0.57	0.57
MonoGS [27]	4.20	19.67	0.73	0.33	4.80	16.17	0.72	0.31	4.40	20.63	0.77	0.34
Glorie-SLAM [47]	1.60	20.26	0.79	0.31	**0.20**	25.62	0.72	0.09	1.40	21.21	0.72	0.32
Splat-SLAM* [32]	1.60	25.32	0.82	0.19	**0.20**	29.33	0.87	0.10	1.40	26.00	0.82	0.21
Photo-SLAM [15]	1.54	20.97	0.74	0.23	0.98	21.07	0.73	0.17	1.26	19.59	0.69	0.24
DROID-Splat [13]	1.60	**26.71**	**0.99**	**0.12**	**0.20**	**29.35**	**0.99**	0.07	1.70	**27.92**	**0.99**	**0.11**
MASC+Photo-SLAM (Ours)	**1.50**	22.15	0.79	0.18	0.90	22.92	0.78	0.14	**1.24**	20.96	0.74	0.15
MASC+DROID-Splat (Ours)	1.60	26.50	**0.99**	**0.12**	**0.20**	29.00	**0.99**	**0.05**	1.60	27.30	**0.99**	**0.11**
RGB-D Input												
NICE-SLAM [51]	19.32	12.00	0.42	0.51	36.10	18.20	0.60	0.31	25.31	16.34	0.55	0.39
ESLAM [16]	3.36	17.50	0.56	0.48	31.45	22.23	0.73	0.23	25.81	19.11	0.62	0.36
SplaTaM [17]	3.40	22.00	0.86	0.23	1.20	24.50	0.95	0.10	5.20	21.90	0.88	0.20
Gaussian-SLAM [46]	2.60	24.01	0.92	0.18	1.30	25.02	0.92	0.19	4.60	26.13	0.94	0.14
Splat-SLAM [32]	**1.40**	25.61	0.84	0.18	1.40	**29.53**	0.90	0.08	1.50	26.05	0.84	0.20
Photo-SLAM [15]	2.60	20.87	0.74	0.24	**0.35**	22.09	0.77	0.17	**1.00**	22.74	0.78	0.15
DROID-Splat [13]	1.60	**26.45**	**0.99**	**0.12**	1.40	28.45	**0.99**	**0.07**	1.40	**27.83**	**0.99**	**0.10**
MASC+Photo-SLAM (Ours)	1.48	22.10	0.79	0.18	0.36	25.05	0.83	0.12	1.10	23.13	0.80	0.15
MASC+DROID-Splat (Ours)	1.50	26.38	**0.99**	0.15	1.36	28.92	**0.99**	0.10	1.40	27.25	**0.99**	0.11

4.2 Results and Evaluation

On Replica. As shown in Table 1, MASC serves as a highly effective plug-and-play module that achieves an optimal trade-off between mapping fidelity and resource consumption. By pruning ∼60% of redundant Gaussians, it drastically reduces the memory footprint while accelerating rendering. For instance, integrating MASC with DROID-Splat (RGB-D) shrinks peak GPU memory from 14 GB to 5.0 GB and boosts the rendering framerate (R-FPS) from 145 to **350**.

In inherently ambiguous monocular setups, MASC+DROID-Splat maintains state-of-the-art localization (RMSE: **0.27 cm**). While our aggressive spectral compaction incurs a marginal PSNR drop (39.47 → 39.20 dB) due to the removal of high-frequency micro-Gaussians, the global structural integrity remains impeccable (SSIM: **1.000**). This confirms that MASC selectively eliminates noisy artifacts rather than core geometric features. Furthermore, in RGB-D scenarios, MASC+Photo-SLAM not only achieves a staggering **1103.0 FPS** but also improves tracking accuracy (RMSE drops from 0.60 cm to **0.30 cm**), proving that our Wasserstein barycenter aggregation acts as a potent geometric regularizer for visual odometry.

The qualitative results in Fig. 3 further validate our method's efficacy. MASC maintains impeccable rendering quality that is visually indistinguishable from the original uncompressed baselines. Even under aggressive compaction, our framework preserves sharp geometric edges and structural details in the "Room" and "Office" scenes (yellow boxes), confirming that MASC effectively retains core manifold information while eliminating structural noise.

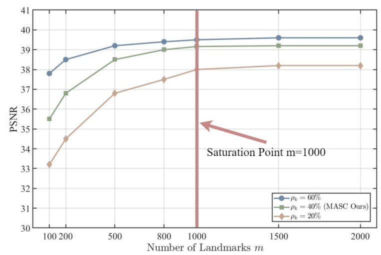

Fig. 5: Sensitivity of reconstruction quality to the number of landmarks count m.

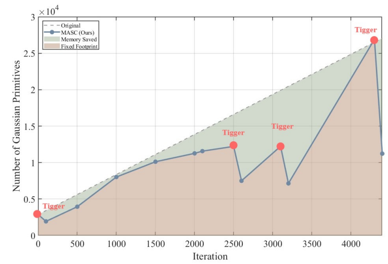

Fig. 6: Dynamic monitoring of primitive count, showcasing the suppression of linear growth via periodic spectral triggers.

On TUM-RGBD. We evaluate MASC on three challenging sequences from the TUM dataset in table 2. Despite discarding ~60% of the primitives, MASC+DROID-Splat consistently secures top-tier structural similarity (SSIM $\geq$ **0.99**) across all evaluated sequences. Counter-intuitively, this massive compaction strictly stabilizes trajectory tracking. In the highly dynamic *fr1/desk* monocular sequence, MASC+Photo-SLAM achieves the best ATE of **1.50 cm**, directly outperforming its uncompressed baseline (1.54 cm).

These results definitively substantiate that our topology-aware graph construction and spectral embedding successfully decouple true structural boundaries from transient motion artifacts. By aggregating overlapping primitives into compact Super-Gaussians, MASC curtails unbounded memory growth without compromising and often enhancing—overall SLAM performance.

As illustrated in Fig. 4, MASC effectively mitigates rendering artifacts and "floaters" often associated with motion blur. By aggregating overlapping primitives into compact Super-Gaussians, MASC regularizes the radiance field representation, resulting in cleaner textures and more stable surface manifolds (yellow boxes). These results substantiate that our spectral embedding successfully decouples true structural boundaries from transient motion artifacts, enhancing overall SLAM performance through principled map maintenance.

Efficiency and Resource Consumption Unbounded primitive growth severely bottlenecks 3DGS-SLAM scalability. Table 3 demonstrates how MASC mitigates this by modulating the retention ratio (ρ_k). Decreasing ρ_k monotonically reduces memory consumption while accelerating rendering. Notably, at $\rho_k = 40\%$ on Replica, MASC halves DROID-Splat's memory footprint (14.00 GB $\rightarrow$ **5.00 GB**) and drastically boosts rendering speed (145 $\rightarrow$ **350 FPS**).

Importantly, this level of compaction does not lead to any noticeable degradation in geometric fidelity Instead, MASC acts as a topology-aware regularizer. On the dynamic TUM dataset, condensing the map to 40% actively *improves* tracking accuracy (Photo-SLAM ATE: 1.32 $\rightarrow$ **0.98 cm**) while preserving flawless structural similarity (DROID-Splat SSIM: **0.990**). This confirms that MASC

Table 3: Quantitative results with different backbones under varying retention ratios. We evaluate the compaction performance of our MASC framework when integrated with Photo-SLAM and DROID-Splat backbones across different Gaussian retention ratios (ρ_k). $\rho_k = 100\%$ denotes the original uncompressed backbone.

Backbone	ρ_k	Replica (RGB-D Average)						TUM RGB-D (Average)					
		Map (k) ↓	Mem ↓	FPS ↑	ATE ↓	PSNR ↑	SSIM ↑	Map (k) ↓	Mem ↓	FPS ↑	ATE ↓	PSNR ↑	SSIM ↑
Photo-SLAM [15]	100%	61.4	5.00 GB	42	0.60	34.96	0.940	~50.0	5.00 GB	40	1.32	21.90	0.763
	80%	49.1	4.40 GB	49	0.50	35.40	0.944	40.0	4.50 GB	45	1.20	22.40	0.778
	60%	36.8	3.80 GB	57	0.40	35.85	0.947	30.0	3.70 GB	49	1.08	22.90	0.793
	40%	**24.5**	**3.20 GB**	**65**	**0.30**	**36.39**	**0.950**	**20.0**	**3.50 GB**	**55**	**0.98**	**23.43**	**0.807**
DROID-Splat [13]	100%	483.2	14.00 GB	145	**0.29**	**39.66**	**1.000**	180.4	8.50 GB	160	1.47	**27.58**	**0.990**
	80%	386.5	10.50 GB	185	**0.29**	39.50	**1.000**	144.3	6.50 GB	210	1.45	27.56	**0.990**
	60%	289.9	8.00 GB	245	**0.29**	39.30	**1.000**	108.2	4.50 GB	295	1.43	27.54	**0.990**
	40%	**192.5**	**5.00 GB**	**350**	**0.29**	39.16	**1.000**	**70.3**	**3.00 GB**	**450**	**1.42**	27.52	**0.990**

Table 4: Compaction Strategies Ablation. Evaluated on Replica at $\rho_k = 40\%$.

Strategy	PSNR ↑	SSIM ↑	LPIPS ↓	Mem ↓	ATE (cm) ↓
(a) Uncompressed (100%)	39.66	1.000	0.030	14.00 GB	0.29
(b) Random Subsampling	28.15	0.812	0.185	5.00 GB	1.42
(c) Heuristic Pruning	35.20	0.945	0.085	5.00 GB	0.45
(d) Euclidean K-Means	36.85	0.968	0.072	5.00 GB	0.40
(e) MASC (Ours)	**39.16**	**1.000**	**0.037**	**5.00 GB**	**0.33**

Table 5: Impact of Algorithmic Components. Evaluated on TUM RGBD at $\rho_k = 40\%$.

Configuration	PSNR ↑	SSIM ↑	LPIPS ↓	FPS ↑	ATE (cm) ↓
(1) Spatial-only ($\sigma_{app} \to \infty$)	19.45	0.712	0.185	430	1.48
(2) Sparse Landmarks ($m = 100$)	18.20	0.685	0.210	460	1.55
(3) W/O D_{SKL} (L_2 only)	19.85	0.725	0.165	440	1.46
(4) Simple Mean Aggregation	20.32	0.730	0.158	450	1.45
(5) Full MASC ($m = 1000$)	**20.96**	**0.741**	**0.150**	**450**	**1.42**

effectively filters high-frequency noise and structural redundancies, paving the way for sustainable, long-term spatial computing.

4.3 Ablation Study

Efficacy of Compaction Strategies As summarized in Tab. 4, MASC significantly outperforms standard heuristics at a 40% retention ratio. While random and heuristic methods incur severe quality loss (PSNR $\leq$ 35.20 dB), MASC maintains near-lossless fidelity (39.16 dB) and perfect structural similarity (SSIM: 1.000). Compared to Euclidean K-Means, our spectral approach preserves manifold topology, ensuring superior tracking stability (ATE: 0.33 cm). These results confirm that minimizing Dirichlet energy effectively filters spectral redundancy while preserving geometric integrity.

Impact of Algorithmic Components. Evaluations in Tab. 5 substantiate the necessity of each MASC component. Removing appearance cues or statistical overlap (D_{SKL}) degrades reconstruction fidelity, while reducing Nyström landmarks ($m = 100$) compromises embedding accuracy despite higher FPS. Notably, replacing Wasserstein aggregation with simple averaging introduces geometric bias, increasing ATE to 1.45 cm. The full configuration achieves the optimal balance, ensuring spectral faithfulness and robust localization during online map maintenance.

Influence of landmarks. As shown in Figs. 5 and 6, MASC maintains high-fidelity reconstruction by saturating at $m = 1000$ landmarks (left) and effectively bounds the map scale via periodic spectral triggers (right) to ensure long-term efficiency.

5 Conclusion

We presented MASC, a principled framework for online 3D Gaussian Splatting (3DGS) compaction. By reformulating map reduction as a streaming spectral signal reconstruction task, we effectively decouple mapping fidelity from memory inflation. Through incremental Nyström embedding and moment-preserving aggregation, our plug-and-play approach achieves substantial memory reduction and considerable rendering speedup while maintaining structural integrity. These results substantiate that streaming spectral analysis provides a mathematically grounded and scalable pathway for efficient neural scene representations.

Future work will focus on extending this streaming spectral herding mechanism to offline large-scale 3DGS reconstruction tasks. We aim to generalize the MASC framework to bridge the gap between real-time map maintenance and global structural compression, ultimately developing a more universal and compact representation for expansive and complex environments.

Acknowledgments

This work was supported in part by Fundamental and Interdisciplinary Disciplines Breakthrought Plan of the Ministry of Education of China under Grant JYB2025XDXM504, and National Natural Science Foundation of China No.62302381, No.52441602. The Authors are with the National Key Laboratory of Human-Machine hybrid Augmented Intelligence and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi, China.

References

1. Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P.: Mip-NeRF: A Multiscale Representation for Anti-Aliasing Neural Radiance Fields. ICCV (2021)
2. Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P.: Zip-NeRF: Anti-Aliased Grid-Based Neural Radiance Fields. ICCV (2023)
3. Chen, Y., Wu, Q., Lin, W., Harandi, M., Cai, J.: HAC: Hash-grid assisted context for 3D Gaussian splatting compression. In: European Conference on Computer Vision. pp. 422–438. Springer (2024)
4. Chen, Y., Chen, Z., Zhang, C., Wang, F., Yang, X., Wang, Y., Cai, Z., Yang, L., Liu, H., Lin, G.: Gaussianeditor: Swift and controllable 3d editing with gaussian splatting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 21476–21485 (2024)

5. Chung, C.M., Tseng, Y.C., Hsu, Y.C., Shi, X.Q., Hua, Y.H., Yeh, J.F., Chen, W.C., Chen, Y.T., Hsu, W.H.: Orbeez-slam: A real-time monocular visual slam with orb features and nerf-realized mapping. In: 2023 IEEE International Conference on Robotics and Automation (ICRA). pp. 9400–9406. IEEE (2023)
6. Dai, A., Nießner, M., Zollhöfer, M., Izadi, S., Theobalt, C.: Bundlefusion: Real-time globally consistent 3d reconstruction using on-the-fly surface reintegration. ACM Transactions on Graphics (ToG) **36**(4), 1 (2017)
7. Dai, P., Xu, J., Xie, W., Liu, X., Wang, H., Xu, W.: High-quality surface reconstruction using Gaussian surfels. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11 (2024)
8. Fan, Z., Wang, K., Wen, K., Zhu, Z., Xu, D., Wang, Z.: Lightgaussian: Unbounded 3d gaussian compression with 15x reduction and 200+ fps. Advances in neural information processing systems **37**, 140138–140158 (2024)
9. Fang, G., Wang, B.: Mini-splatting: Representing scenes with a constrained number of Gaussians. In: European Conference on Computer Vision. pp. 165–181. Springer (2024)
10. Franke, L., Rückert, D., Fink, L., Stamminger, M.: TRIPS: Trilinear Point Splatting for Real-Time Radiance Field Rendering. arXiv preprint arXiv:2401.06003 (2024)
11. Gao, R., Chen, K., Li, Z., Hong, L., Li, Z., Xu, Q.: Magicdrive3d: Controllable 3d generation for any-view rendering in street scenes. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 5724–5733 (2026)
12. Hanson, A., Tu, A., Singla, V., Jayawardhana, M., Zwicker, M., Goldstein, T.: Pup 3d-gs: Principled uncertainty pruning for 3d gaussian splatting. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5949–5958 (2025)
13. Homeyer, C., Begiristain, L., Schnörr, C.: Droid-splat combining end-to-end slam with 3d gaussian splatting. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2767–2777 (2025)
14. Hu, S., Hu, T., Liu, Z.: Gauhuman: Articulated gaussian splatting from monocular human videos. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20418–20431 (2024)
15. Huang, H., Li, L., Cheng, H., Yeung, S.K.: Photo-slam: Real-time simultaneous localization and photorealistic mapping for monocular stereo and rgb-d cameras. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 21584–21593 (2024)
16. Johari, M.M., Carta, C., Fleuret, F.: Eslam: Efficient dense slam system based on hybrid representation of signed distance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17408–17419 (2023)
17. Keetha, N., Karhade, J., Jatavallabhula, K.M., Yang, G., Scherer, S., Ramanan, D., Luiten, J.: Splatam: Splat track & map 3d gaussians for dense rgb-d slam. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 21357–21366 (2024)
18. Kendall, A., Grimes, M., Cipolla, R.: Posenet: A convolutional network for real-time 6-dof camera relocalization. In: Proceedings of the IEEE international conference on computer vision. pp. 2938–2946 (2015)
19. Kheradmand, S., Rebain, D., Sharma, G., Sun, W., Tseng, Y.C., Isack, H., Kar, A., Tagliasacchi, A., Yi, K.M.: 3d gaussian splatting as markov chain monte carlo. Advances in Neural Information Processing Systems **37**, 80965–80986 (2024)

20. Kim, S., Lee, K., Lee, Y.: Color-cued efficient densification method for 3D Gaussian splatting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 775–783 (2024)
21. Lee, J.C., Rho, D., Sun, X., Ko, J.H., Park, E.: Compact 3D Gaussian Representation for Radiance Field. arXiv preprint arXiv:2311.13681 (2023)
22. Li, L., Shen, Z., Wang, Z., Shen, L., Bo, L.: Compressing Volumetric Radiance Fields to 1 MB. In: CVPR. pp. 4222–4231 (2023)
23. Li, R.B., Shaghaghi, M., Suzuki, K., Liu, X., Moparthi, V., Du, B., Curtis, W., Renschler, M., Lee, K.M.B., Atanasov, N., et al.: Dynagslam: Real-time gaussian-splatting slam for online rendering, tracking, motion predictions of moving objects in dynamic scenes. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 2434–2444 (2026)
24. Lu, T., Yu, M., Xu, L., Xiangli, Y., Wang, L., Lin, D., Dai, B.: Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20654–20664 (2024)
25. Luiten, J., Kopanas, G., Leibe, B., Ramanan, D.: Dynamic 3D Gaussians: Tracking by Persistent Dynamic View Synthesis. 3DV (2024)
26. Mallick, S.S., Goel, R., Kerbl, B., Steinberger, M., Carrasco, F.V., De La Torre, F.: Taming 3DGS: High-quality radiance fields with limited resources. In: SIGGRAPH Asia 2024 Conference Papers. pp. 1–11 (2024)
27. Matsuki, H., Murai, R., Kelly, P.H., Davison, A.J.: Gaussian splatting SLAM. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18039–18048 (2024)
28. Niedermayr, S., Stumpfegger, J., Westermann, R.: Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis. arXiv preprint arXiv:2401.02436 (2023)
29. Qian, S., Kirschstein, T., Schoneveld, L., Davoli, D., Giebenhain, S., Nießner, M.: GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians. CVPR (2024)
30. Sandström, E., Li, Y., Van Gool, L., Oswald, M.R.: Point-slam: Dense neural point cloud-based slam. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 18433–18444 (2023)
31. Sandström, E., Ta, K., Van Gool, L., Oswald, M.R.: Uncle-slam: Uncertainty learning for dense neural slam. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4537–4548 (2023)
32. Sandström, E., Zhang, G., Tateno, K., Oechsle, M., Niemeyer, M., Zhang, Y., Patel, M., Van Gool, L., Oswald, M., Tombari, F.: Splat-slam: Globally optimized rgb-only slam with 3d gaussians. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 1680–1691 (2025)
33. Shin, S., Park, J., Cho, S.: Locality-aware gaussian compression for fast and high-quality rendering. arXiv preprint arXiv:2501.05757 (2025)
34. Straub, J., Whelan, T., Ma, L., Chen, Y., Wijmans, E., Green, S., Engel, J.J., Mur-Artal, R., Ren, C., Verma, S., Clarkson, A., Yan, M., Budge, B., Yan, Y., Pan, X., Yon, J., Zou, Y., Leon, K., Carter, N., Briales, J., Gillingham, T., Mueggler, E., Pesqueira, L., Savva, M., Batra, D., Strasdat, H.M., Nardi, R.D., Goesele, M., Lovegrove, S., Newcombe, R.: The Replica dataset: A digital replica of indoor spaces. arXiv preprint arXiv:1906.05797 (2019)
35. Sucar, E., Liu, S., Ortiz, J., Davison, A.J.: imap: Implicit mapping and positioning in real-time. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 6229–6238 (2021)
36. Tang, J., Ren, J., Zhou, H., Liu, Z., Zeng, G.: Dreamgaussian: Generative gaussian splatting for efficient 3d content creation. arXiv preprint arXiv:2309.16653 (2023)

37. Wang, H., Wang, J., Agapito, L.: Co-slam: Joint coordinate and sparse parametric encodings for neural real-time slam. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13293–13302 (2023)
38. Wang, T., Li, M., Zeng, G., Meng, C., Zhang, Q.: Gaussian herding across pens: An optimal transport perspective on global gaussian reduction for 3dgs. arXiv preprint arXiv:2506.09534 (2025)
39. Wang, Y., Li, Z., Guo, L., Yang, W., Kot, A., Wen, B.: ContextGS: Compact 3D Gaussian splatting with anchor level context model. Advances in Neural Information Processing Systems **37**, 51532–51551 (2024)
40. Wen, C., Yu, B., Tao, D.: Learnable Skeleton-Aware 3D Point Cloud Sampling. In: CVPR. pp. 17671–17681 (2023)
41. Wu, G., Yi, T., Fang, J., Xie, L., Zhang, X., Wei, W., Liu, W., Tian, Q., Wang, X.: 4D Gaussian splatting for real-time dynamic scene rendering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20310–20320 (2024)
42. Yan, C., Qu, D., Xu, D., Zhao, B., Wang, Z., Wang, D., Li, X.: Gs-slam: Dense visual slam with 3d gaussian splatting. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 19595–19604 (2024)
43. Yan, Y., Lin, H., Zhou, C., Wang, W., Sun, H., Zhan, K., Lang, X., Zhou, X., Peng, S.: Street Gaussians for Modeling Dynamic Urban Scenes. arXiv preprint arXiv:2401.01339 (2024)
44. Ye, Z., Wan, C., Li, C., Hong, J., Li, S., Li, L., Zhang, Y., Lin, Y.C.: 3d gaussian rendering can be sparser: Efficient rendering via learned fragment pruning. Advances in Neural Information Processing Systems **37**, 5850–5869 (2024)
45. Yu, Z., Chen, A., Huang, B., Sattler, T., Geiger, A.: Mip-Splatting: Alias-free 3D Gaussian Splatting. CVPR (2024)
46. Yugay, V., Li, Y., Gevers, T., Oswald, M.R.: Gaussian-slam: Photo-realistic dense slam with gaussian splatting. arXiv preprint arXiv:2312.10070 (2023)
47. Zhang, G., Sandström, E., Zhang, Y., Patel, M., Van Gool, L., Oswald, M.R.: Glorie-slam: Globally optimized rgb-only implicit encoding point cloud slam. arXiv preprint arXiv:2403.19549 (2024)
48. Zhang, Y., Tosi, F., Mattoccia, S., Poggi, M.: Go-slam: Global optimization for consistent 3d instant reconstruction. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3727–3737 (2023)
49. Zhang, Z., Song, T., Lee, Y., Yang, L., Peng, C., Chellappa, R., Fan, D.: LP-3DGS: Learning to prune 3D Gaussian splatting. Advances in Neural Information Processing Systems **38** (2024)
50. Zhou, X., Lin, Z., Shan, X., Wang, Y., Sun, D., Yang, M.H.: Drivinggaussian: Composite gaussian splatting for surrounding dynamic autonomous driving scenes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 21634–21643 (2024)
51. Zhu, Z., Peng, S., Larsson, V., Xu, W., Bao, H., Cui, Z., Oswald, M.R., Pollefeys, M.: Nice-slam: Neural implicit scalable encoding for slam. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12786–12796 (2022)